



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets



Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00127581.7

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE,
LA HAYE, LE

31/05/01

This Page Blank (uspto)



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

**Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation**

Anmeldung Nr.:
Application no.: 00127581.7
Demande n°:

Anmeldetag:
Date of filing: 15/12/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
International Business Machines Corporation
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:

Method and system for content off-loading in a document processing system using stub documents

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragsstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du dépôt:

Bemerkungen:
Remarks:
Remarques:

This Page Blank (uspto)

15-12-2000

EP00127581.7

SPEC

DE920000113

- 1 -

D E S C R I P T I O N

**Method and System for Content Off-Loading in a Document
Processing System using Stub Documents**Background of the Invention

The invention relates to data processing environments with large document repositories and, more specifically, to a method and system for handling content off-loading from a document processing system to a remote repository.

Known mailing client applications like Lotus[™] Notes[™] or Microsoft[™] Outlook[™] contain continuously growing document repositories, namely the incoming and outgoing notes or emails often including large attachments like text documents, graphics or even storage consuming digitized pictures. Therefore, e.g. a Lotus Notes application uses a Lotus Domino[™] database from which a tool like IBM Content Manager CommonStore[™] for Lotus Domino (CSLD) is used to move documents stored in that database to an archive physically located on a different device like a tape storage. CSLD thereupon allows to access documents that have previously been archived.

CSLD also allows to access documents that have been archived from any archive client application (e.g. scanning applications, CommonStore for SAP[™], etc). When documents are retrieved from the archive to a Notes database, a Lotus Notes document is created.

IBM Content Manager CommonStore[™] for Lotus Domino (CSLD) is an according tool to move Lotus Notes documents in various formats to an archive. CSLD also allows to access documents that have been archived from any archive client application (e.g. scanning applications, CommonStore for SAP[™], etc). When documents are

15-12-2000

EP00127581.7

SPEC

DE920000113

- 2 -

retrieved from the archive to a Notes database, a Lotus Notes document is created.

The IBM Archive Content Manager, and another tool called "OnDemand", maintain an index about archived documents. This means that archived documents can be deleted from Lotus Notes, since it is possible to find them later by searching the archive's index. In contrast, Tivoli Storage Manager (TSM) does not provide an index by its own, but rather leaves it up to the archiving application to maintain an index. That is, TSM itself does not allow to search for archived documents.

CSLD uses the original documents within Notes to maintain the index: When a Notes document is archived via CSLD, it is assigned a unique identifier (ID) by the archive. CSLD generally writes this document archive ID to a field in the original Notes document. This allows to retrieve an archived document by ID without performing a search in the archive.

A drawback of the above prior art approach is that, when a document is deleted from Notes, the link to the archived document is completely lost. With Content Manager and OnDemand, the archived document could still be retrieved via an archive search. For TSM, however, since it does not provide an index to search over, there is no way to retrieve an archived document once the only Notes document containing the link to it is deleted. Therefore, CSLD does not allow to delete a document from Notes that has been archived to TSM.

However, there is a need also for CSLD to release expensive disk space by archiving/off-loading complete Notes documents.

Summary of the Invention

It is therefore an object of the present invention to provide a

15-12-2000

EP00127581.7

DE920000 SPEC

- 3 -

method and system for handling content off-loading to a large document repository, which are less resource consuming than the prior art approaches.

Another object is to provide a user-friendly mechanism for off-loading and retrieving content.

The above objects are achieved by the features of the independent claims. Advantageous embodiments are subject matter of the subclaims.

The concept of the invention is that a document including any possible attachments is copied to a remote repository and stripped down to a stub document containing at least the information required to retrieve the copied document from the remote repository. During retrieval, the retrieved content is re-inserted into the stub document to restore the original document.

In other words, the invention proposes a document processing on an original document where content is cut-off or separated from the original document and deleted and the complete original document migrated (off-loaded) to the remote repository. In the stripped-down document only information which enables to identify the off-loaded document on the remote repository and to retrieve it from the remote repository is provided. A few descriptive parts of the document are left in the stub document which allow to identify the stub document in the document processing system.

The original document and the stub document have the same document ID in the underlying document processing system. Therefore, although the document has been off-loaded, existing links to it remain still valid.

15-12-2000

EP00127581.7

DE920000

SPEC

- 4 -

It is emphasized that the original document and the stub document are the same document just in two different versions, not being clones, placeholders or even copies.

The proposed mechanism is less resource consuming than the prior art approaches and can advantageously be used in mail clients where mails potentially including attachments are archived on a remote mail server. As a first, storage is released due to the proposed 'down-stripping' of the original documents. Secondly, since the stub documents still contain a few descriptive fields, it is possible to search for off-loaded documents in the document processing system, although the remote repository may not provide a search index or mechanism.

It is understood that the remote repository can be also a local hard disk.

Brief Description of the Drawings

In the following, the present invention is described in more detail by way of embodiments from which further features and advantages of the invention become evident whereby

Fig. 1 is a flow diagram illustrating the various steps to archive a document and create a stub document from it, in accordance with the invention; and

Fig. 2 schematically shows the structure of a Lotus Notes document before and after stripping it down to a stub document according to the invention.

Detailed Description of the Drawings

15-12-2000

EP00127531.7

DE920000

SPEC

- 5 -

Referring to Fig. 1, an archiving request for a Lotus Notes document 101 is issued to IBM Content Manager CommonStore™ for Lotus Domino (CSLD) 102 which copies the document 101 to a remote archive 103, the remote archive 103 being an example for a document repository. After archiving, CSLD 102 creates a stub document 104 from the original document 101 by stripping it down. The original document 101 and the stripped-down document 104 have the same document ID 105.

The size of a stub document is only a small percentage of the original document. In the present example, the stripping-down process reduces the document size of the original Lotus Notes document from 100 kByte to about 1 kByte.

In CSLD, when a document has been archived successfully, it can be converted to a stub automatically and synchronously by applying LotusScript or Java code to it. This code can be customized so that administrators can decide which items to remove from documents.

The above and in the following described mechanism to create stub files is based on the assumption that the document processing system is Lotus Notes. But it is noteworthy that the underlying concept of the invention can be applied also to other document processing environment. Technically, documents in Lotus Notes are basically a collection of items. All content except attachments and OLE objects is kept in items. Therefore, a stub document in that environment is a Notes document from which all large items have been removed. Further, a stub contains the item that contains the link to the archived document.

In addition, the stub document contains just enough information to allow the document to be displayed in a view or folder, and the document's readers fields. For example, a stub of a Notes email (Memo) should contain the sender, receiver list, date/time the mail was sent, the subject and the link to the archived

document. When a stub document is displayed in a view/folder, it cannot be distinguished from regular Notes documents since it contains all items to be displayed in the view/folder.

Now referring to Fig. 2, it is illustrated how a Lotus Notes email document containing an attachment and various other fields, is converted to a stub document in accordance with the invention. The stripping-down process leaves only those fields that are necessary to identify emails among other emails, in the present example the 'Subject', 'Mail Sender', 'Mail Recipients', and the date and time the mail was posted. Also the link to the archived document remains left in the stub document.

In the following it is described in more detail how searching for stub documents and retrieving archived documents are handled according to the invention, in case of a underlying generic document processing system.

Searching for stub documents

As mentioned before, stub documents are regular documents containing a few descriptive fields. Therefore the search mechanism provided by the document processing system, e.g. a full text search, can be used to find even stub documents. Once a stub document is found, the original document can be restored via the archive ID stored in the stub document.

Retrieving archived documents by overwriting stubs

Once a stub is found after a predescribed search, a user can retrieve the corresponding archived (complete) document. CSLD extracts the archive ID from the stub document, and retrieves the document from the archive using the archive ID. Then, the content of the archived document is re-inserted into the stub

15-12-2000

EP00127581.7

SPEC

DE920000115

- 7 -

document. This will restore the original document completely. Even the document's unique ID (UNID) and security properties are preserved.

It is emphasized that the proposed stub creation is not only useful when the above described Tivoli Storage Manager is used as the archive behind CSLD. Even for archives supporting an index you can create stubs from archived documents instead of deleting them after archiving. This allows to search for archived documents in the document processing system, instead of searching in the archive. The search results are returned much faster than searching the archive.

15-12-2000

EP00127581.7

DE920000 SPEC

- 8 -

C L A I M S

1. A method for handling content off-loading from a document processing system to a remote repository, comprising the steps of:

copying a document from the document processing system to the remote repository;

stripping down the original document to a stub document containing at least information enabling to retrieve the document from the remote repository.

2. Method according to claim 1, wherein the stripping down leaves descriptive parts of the document in the stub document in order to identify the stub document in the document processing system.
3. Method according to claim 1 or 2, wherein keeping a link to the document in the repository in the stub document.
4. Method according to any of the preceding claims, wherein during retrieval of a document from the repository, the retrieved content is re-inserted into the stub document to restore the original document.
5. Method according to any of the preceding claims, wherein the stripping down of the document preserves a document's unique identifier thus keeping links to the original document valid.
6. A system for handling content off-loading from a document processing system to a remote repository, comprising the steps of:

means for copying a document from the document processing

15-12-2000

EP00127531.7

DE920000115
SPEC

- 9 -

system to the remote repository;

means for stripping down the original document to a stub document containing at least information enabling to retrieve the document from the remote repository.

7. System according to claim 6, comprising means for re-inserting a retrieved content into the stub document during retrieval of the document from the repository.
8. A data processing program for execution in a data processing system comprising software code portions for performing a method according to any of claims 1 to 5 when said program is run on said computer.
9. A computer program product stored on a computer usable medium, comprising computer readable program means for causing a computer to perform a method according to any of claims 1 to 5 when said program is run on said computer.

15-12-2000

EP00127531.7

SPEC

1 / 2

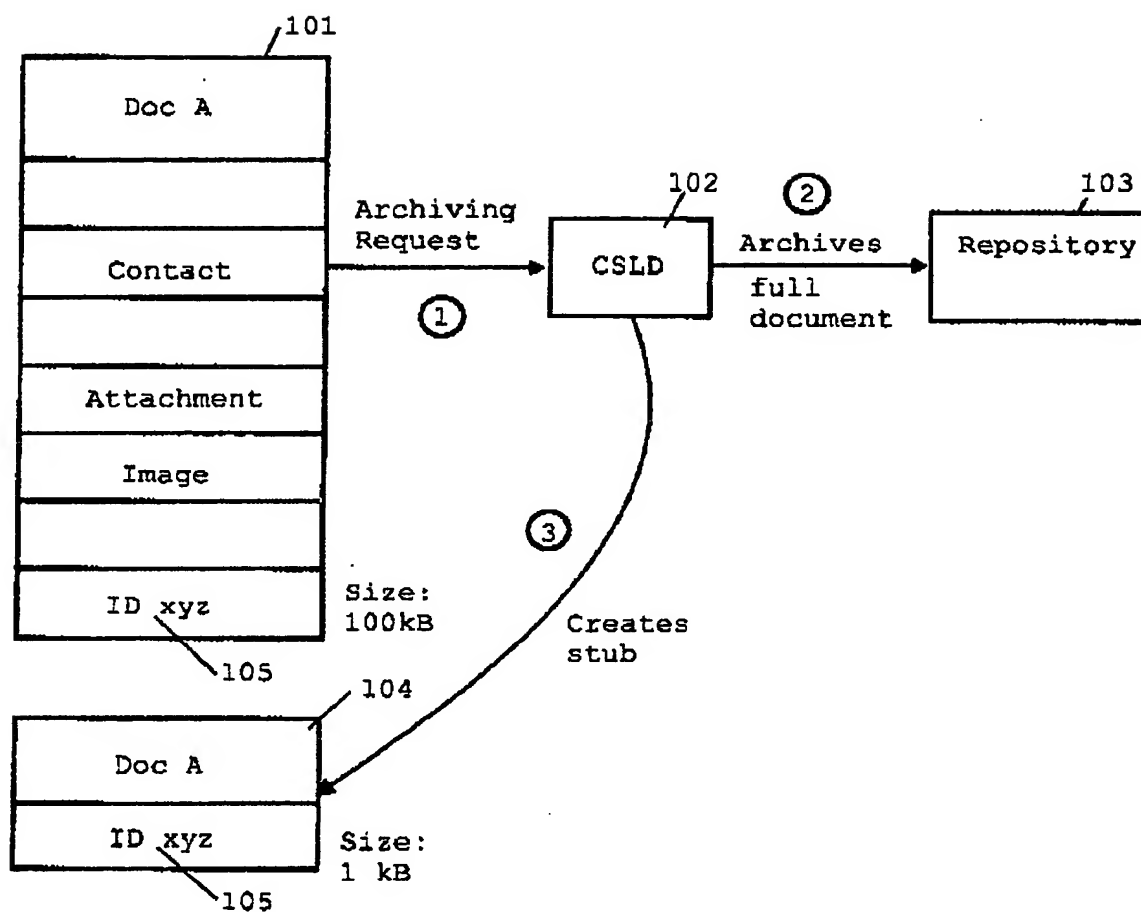


FIG. 1

15-12-2000

EP00127581.7

SPEC

2 / 2

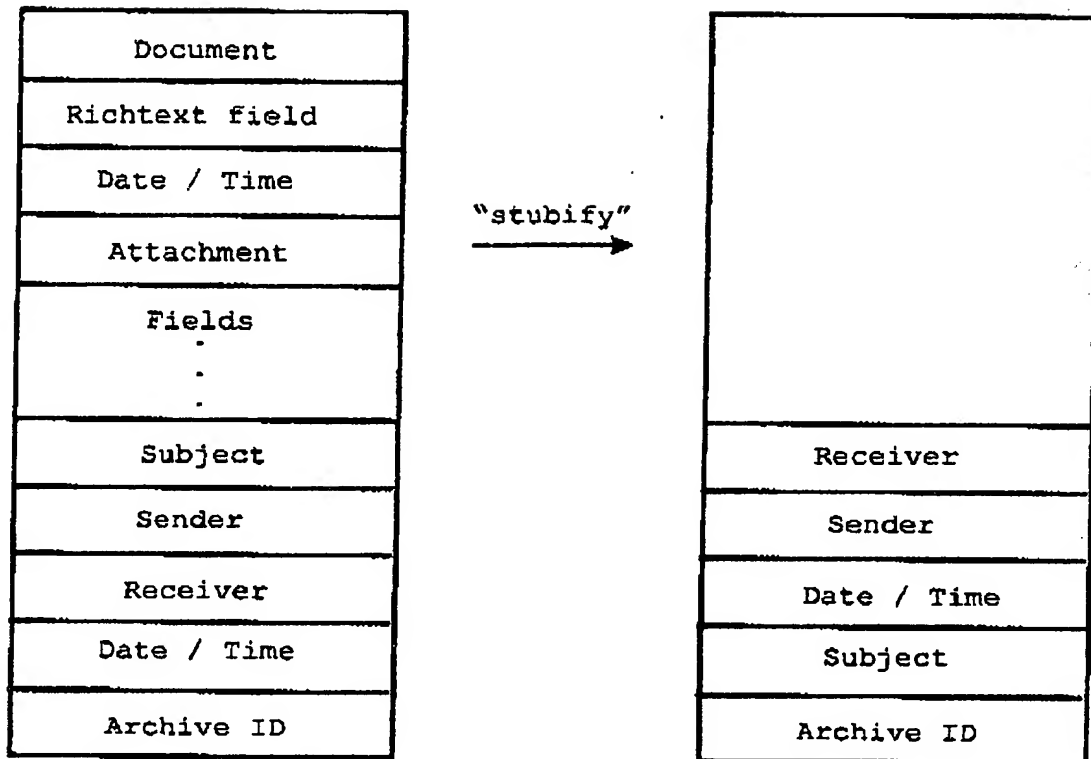


FIG. 2

15-12-2000

EP00127581.7

DE920000113

SPEC

- 1 -

A B S T R A C T

Disclosed are a method and system for handling document or content off-loading from a document processing system to a large repository. The document including any possible attachments is copied to a remote repository and stripped down to a stub document containing at least the information required to retrieve the copied document from the remote repository. During retrieval, the retrieved content is re-inserted into the stub document to restore the original document.

The proposed mechanism is less resource consuming than the prior art approaches and can advantageously be used in mail clients where mails potentially including attachments are archived on a remote mail server.

(Fig. 1)